Probabilistic Records Linkage using Historical Canadian Censuses and Vital Statistics

Stacey A Hallman

PhD Candidate, University of Western Ontario
Research Project with Alain Gagnon, D Ann Herring, and David Earn

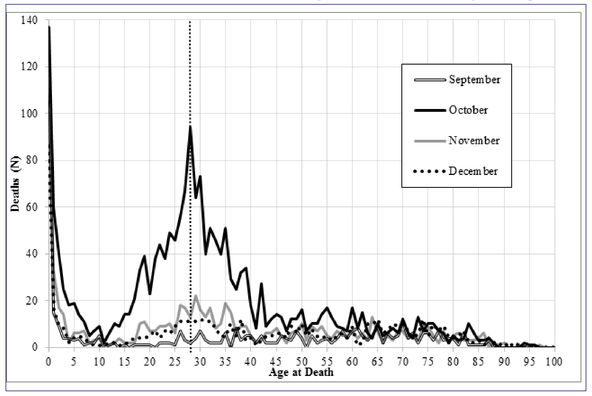
Abstract

This project is involves records linkage to analyze potential mortality risks for young adults in Ontario during the 1918 influenza epidemic. With a team of research assistants at Western University, McMaster University, and the Université de Montréal, death records were linked to birth records in order to determine exact date of birth for the calculation of exact age at death. This date of birth is then compared to age as listed on other historical documents, such as the 1901 and 1911 Canadian censuses. The family environment (or living situation) was transcribed for each individual at each census to examine the impact of socioeconomic conditions throughout the lifecourse on mortality in 1918.

Of the 23,183 deaths registered in Ontario between September and December 1918 and transcribed by the International Infectious Disease Data Archive at McMaster University, 3,316 individuals met the inclusion criteria. Both birth and death must have occurred in Ontario to establish exact date of birth and the individual must have died between the ages of 23 and 35 (born between 1883 and 1895). Of the 3,316 included death records, 2,965 were linked to at least one other record, giving a linkage success rate of 89.4%. This poster analyzes the linkage rates of the death records to birth records from 1883-1895 and the 1901 and 1911 Canadian censuses and uses logistic regression to investigate the important factors that precluded linkage. It evaluates declared age at all three time periods to discuss the suitability of these records for historical demographic analyses of past epidemic disease.

Research Question

- 1918 influenza pandemic:
  - Unexpected increase in young adult mortality.
- Ontario death records show a peak in mortality at age 28.



All-Cause Mortality, Toronto, September to December, 1918.
Source: "Figure 1. Number of deaths by age from all causes in the City of Toronto, September to December, 1918 (September, n=441; October, n=1,885, November n=731, December n=618. Total n=3,675). The vertical line indicates age 28." From Hallman and Gagnon (2014).

- 1890 influenza pandemic occurred 28 years earlier.
  - January to March, 1890.
  - Did early life or *in utero* exposure influence mortality in later life?
- Research to answer the question:

Through a historical demographic lens, are the extant historical records in Ontario suitable for demographic analyses of past infectious disease?

Methods

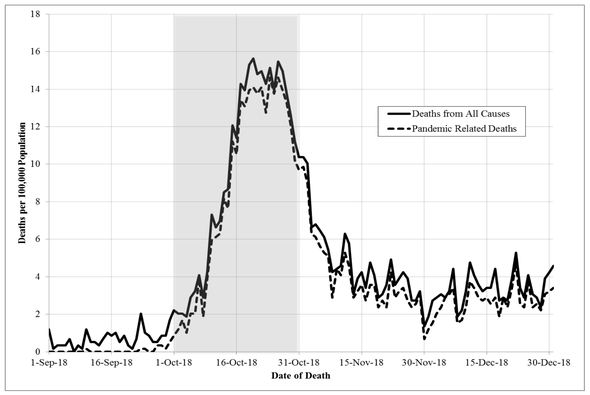
- Interested in young adult mortality:
  - Died between ages 23 and 35.
  - Born between 1883 and 1895.
- Died in Ontario between September and December, 1918:
  - Second (worst) wave of the 1918 pandemic.
  - Birth records linkage: Must be born in Ontario.
  - Excludes Immigrants.
- Age in whole years on death records insufficient to determine potential exposure.

•Created the Western, McMaster, Montreal Influenza Pandemic (WMMIP) database:

- Linking death records provided by the International Infectious Disease Data Archive (IIDDA) at McMaster University to:
  - Birth records.
  - 1901 and 1911 Canadian censuses.
  - Marriage and Attestation papers.

Sample

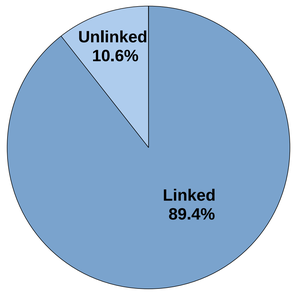
- 23,183 death records for all ages, from September to December, 1918, in Ontario were provided by the IIDDA.
- 3,316 met the inclusion criteria:
  - 1,743 male (52.6%) and 1,573 (47.4%) female.
  - Declared age at death range: 22 to 37.
  - Reconstructed age at death range: 23-35.
  - Pandemic related deaths: 2,816 (84.9%).



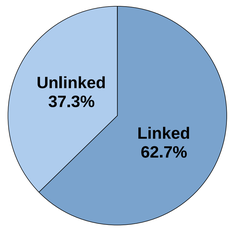
Death Rates from All Causes and Pandemic Related Deaths, Ontario, September to December 1918
Note: For those in the final WMMIP database (N=3,313) and pandemic related deaths (N=2,814) for those with both a month and date of death. Rates calculated by the total number of people in the interpolated population between the ages of 23 and 35 (N= 588,191.22). Highlighted section is October 1918

Linkage Success

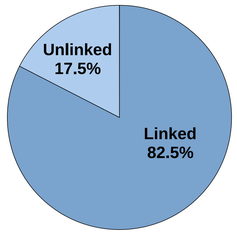
- Linked to *at least* one other record:



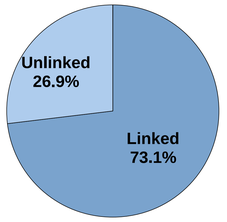
- To Birth Records:



To the 1901 Census:



- To the 1911 Census:



Reconstructed Age at Death

- All linked records compared to form a Reconstructed Date of Birth and Reconstructed Age at the Census
- Accuracy of declared age on the death records:
  - Evidence of age heaping.
  - If inaccurate, most often declared to be 1 year younger.

Reconstructed Age at Death	2+ Years Younger	1 Year Younger	Properly Classified %	1 Year Older	2+ Years Older
23			75.2	16.1	8.7
24	0.5	17.1	65.0	12.9	4.6
25	3.3	13.6	65.8	13.6	3.8
26	5.9	18.6	62.1	9.7	3.7
27	4.9	18.6	62.7	12.2	1.5
28	7.7	18.6	59.6	11.6	2.5
29	8.3	23.0	58.7	8.3	1.6
30	8.8	16.5	64.8	6.5	3.4
31	6.4	22.1	54.4	11.3	5.9
32	9.0	12.9	58.6	13.3	6.2
33	10.1	10.6	63.3	10.3	5.3
34	6.6	17.8	60.4	15.2	
35	10.1	18.9	71.1		

Unlinked Individuals

- Using logistic regression, unlinked individuals tended to be:
  - Aboriginal.
  - From Northern Ontario.
  - More likely to have missing information:
    - Father's Name.
    - Mother's Name.
  - Had possibly itinerant occupations.
  - Older.
  - Died in an institution.
    - Epidemic Underreporting.
- No difference by:
  - Cause of death.
  - Sex.

Key Findings and Conclusions

- Highly successful linkage rate.
- Date of birth best taken from the birth records, but they are incomplete. Second best is the 1901 census. Age in whole years is not useful in either the 1901 or the 1911 census.
- Some results might have been artifacts of linkage process itself.

ACKNOWLEDGEMENTS

- Alain Gagnon and Bill Avison.
- Ann Herring, David Earn, and the IIDDA.
- University of Western Ontario, McMaster University, and the Université de Montréal.
- SSHRC and CIHR.
- Research Assistants:

Enrique Acosta	Jennifer Brooks	Vincent Lortie
Marilyn Amorevieta-Gentil	Michaela Colville	Alex Rewegan
Sophie-Andrée Piché	Astrid Flénon	Sarah Wahoush
Kandace Bogaert	Julie Lacroix	Melissa Yan

REFERENCES

Gagnon, A., M. S. Miller, S. Hallman, R. Bourbeau, D. A. Herring, D. J. D. Earn, and J. Madrenas (2013). Age-Specific Mortality during the 1918 Influenza Pandemic: Unraveling the Mystery of High Young Adult Mortality. PLoS ONE 8(8): e69586

Hallman, Stacey and Alain Gagnon (2014). Does Exposure to Influenza Very Early in Life Affect Mortality Risk During a Subsequent Outbreak? The 1890 and 1918 Pandemics in Canada. *In* Are Modern Environments Bad for Human Health? Revisiting the Second Epidemiological Transition, Molly K. Zuckerman, ed. Pp. 123-138. Toronto: Wiley-Blackwell.